**QR码安全**

Peter Kieseberg, Manuel Leithner, Martin Mulazzani, Lindsay Munroe, Sebastian

Schrittwieser, Mayank Sinha, Edgar Weippl

SBA Research

Favoritenstrasse 16

AT-1040 Vienna, Austria

[1stletterfirstname][lastname]@sba-research.org

**摘要**

本文考察了QR码及其如何用于攻击人类交互和自动化系统。 由于编码信息仅用于机器可读，因此人们无法区分有效和恶意操作的QR码。 虽然人类可能会因为网络钓鱼攻击而下降，但自动读取器最有可能易受SQL注入和命令注入的攻击。 我们的贡献包括对QR码作为攻击矢量的分析，显示攻击者的不同攻击策略和探索其可能的后果。

**1. 介绍**

QR（“快速响应”）码是由日本公司Denso Wave发明的二维条形码。信息在垂直方向和水平方向两方向同时编码，因此比传统的条形码数据多达数百倍（图1）。通过使用相机（例如，内置于智能手机中）捕获代码的照片并利用QR读取器处理图像来访问数据。

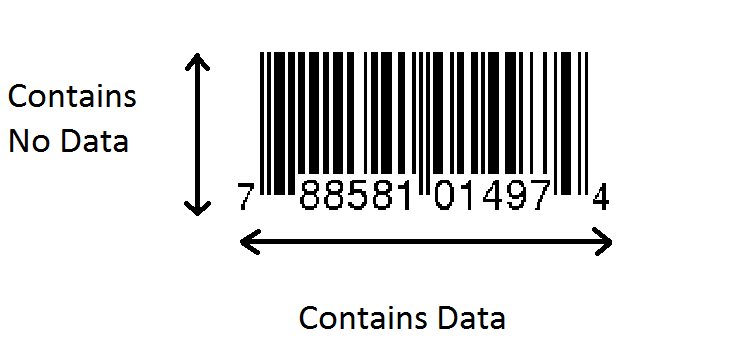


图1：条码

QR码（图2）迅速获得国际知名度的普及，特别是在日本，默认编码汉字符号的能力尤其适合。热门用途包括在海报，招牌，名片，公共交通工具等上存储URL，地址和各种形式的数据。事实上，这种机制具有广泛的潜在应用[4，1，2，13，9]。 例如，运动品牌Umbro已将QR码嵌入到英格兰足球衫领上，将球迷送到一个可以获得奖品的秘密网站。

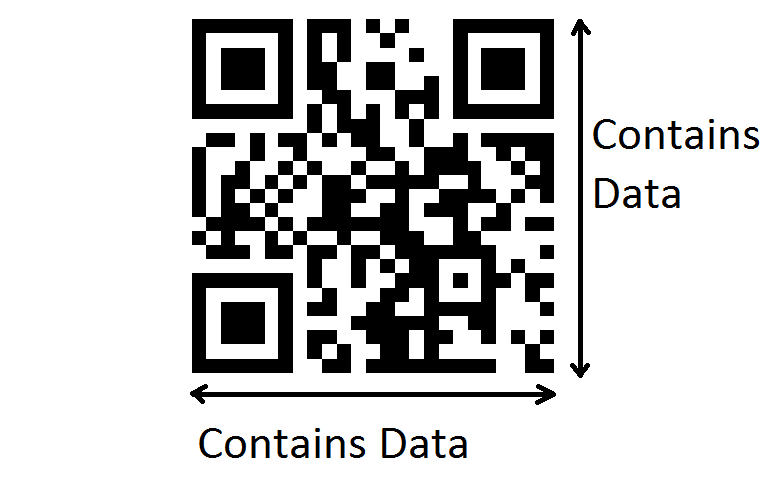


图2：QR码

在本文中，我们将探讨QR码的结构和创建过程以及针对或使用QR码的潜在攻击。 我们概述了纠错能力和可能的方式来更改纠错数据和有效载荷，以便将信息修改或注入现有代码。此外，我们探索了许多可能使攻击者利用用户对嵌入在代码中的内容的信任或处理此类代码的自动过程的许多向量。

我们的主要贡献是：

•概述对QR码不同部分的可能修改，如纠错码或屏蔽，

•描述针对人类（例如网络钓鱼攻击）和自动化过程（例如SQL注入）的最终 攻击媒介。

**2. 背景**

QR码[11]已经超越了一些地区流行的古典条形码。 这在许多情况下源自一个典型的条形码只能容纳20位数字，而QR码最多可以容纳7,089个字符。 结合提供的多样性和可扩展性，这使得QR码的使用比条形码的吸引力更大。统计上，QR码可以在传统条形码的大约十分之一的空间内编码相同数量的数据。QR码的一大特点是它们不需要从一个特定的角度进行扫描，因为QR码可以被读取，而不考虑它们的定位。QR码扫描仪能够确定正确的方式来解码图像，因为位于符号和对齐块的角落中的三个特定方块。

车辆制造商最初使用QR码跟踪零件。过了一段时间，公司开始看到QR码的不同用例。QR码的最受欢迎的商业用途是在电信行业，智能手机越来越多的采用似乎是其流行的最大推动力[13,5,6]。 随着手机技术的不断发展，特别是在移动互联网领域，QR码似乎是快速高效地向用户传达URL的工具。这也允许离线媒体，如杂志，报纸，名片，公共交通工具，标志，T恤或任何其他媒体，可以接受QR码的打印作为在线产品广告的载体[7]。

**2.1 QR码**

QR码由为特定目的而保留的不同区域组成。 在下文中，我们参考QR码的版本2（图3），因为版本1不包含所有模式。

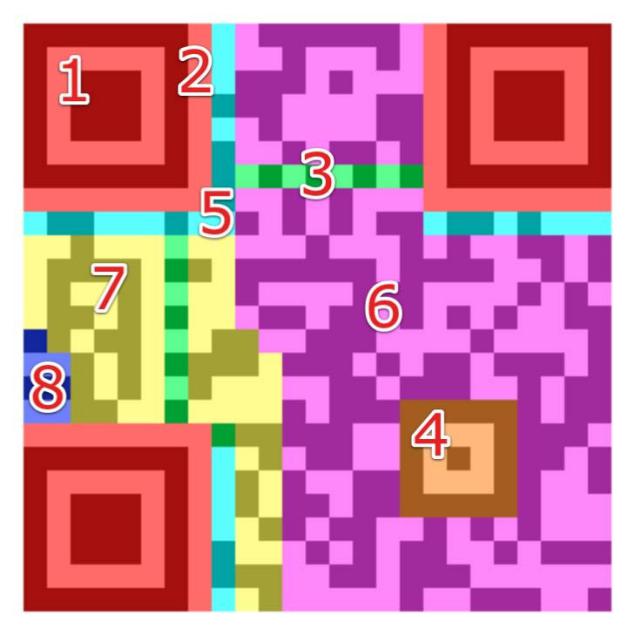


图3：QR码版本2的结构

•Finder图案（1）：取景器图案由位于QR码所有角落的三个相同的结构组成，除了右下角。 每个图案都是基于由黑色模块四周的白色模块所组成的3x3矩阵黑色模块。 Finder图案使解码器软件能够识别QR码并确定正确的方向。

•分隔符（2）：白色分隔符的宽度为一个像素，并提高了Finder 图案的可识别性，因为它们与实际数据分离。

•定时图案（3）：定时图案中交替的黑白模块使解码器软件能够确定单个模块的宽度。

•对齐图案（4）：对齐图案支持解码软件补偿中度图像失真。版本1的 QR码没有对齐图案。随着代码的增加，更多的对齐图案会被添加。

•格式信息（5）：形成信息部分由分隔符旁边的15位组成，并存储有关QR码的纠错级别和所选屏蔽图案的信息。

•数据（6）：将数据转换为比特流，然后存储在数据部分的8位部分（称为码字）中。

•纠错（7）：类似于数据部分，纠错码存储在纠错部分的8位长码字中。

•剩余位（8）：如果数据和纠错位不能被划分为8位码字而无余数，则此部分由空位组成。

整个QR码必须被所谓的“安静区”包围，这个区域与白色模块相同的色调，以改善解码器软件的代码识别。

**2.2 容量和纠错码**

QR码的容量取决于几个因素。 除了定义其大小（模块数量）的代码版本之外，所选择的纠错级别和编码数据的类型也影响容量。

•版本：40个不同版本的QR码主要有不同的模块数量。版本1由21x21模块组成，最多133个（最低纠错级别），可用于存储编码数据。最大的QR码（版本40）的大小为177x177模块，最多可存储23,648个数据模块。

•纠错级别：QR码中的纠错是基于Reed-Solomon码[14]，一种特定形式的BCH纠错码[3,8]。错误纠正有四个级别（表1），用户可以在创建时选择。

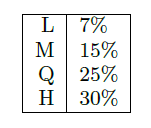


表1：纠错级别

更高的纠错水平增加了用于纠错的码字的百分比，从而减少了可以存储在码内的数据量。

•编码数据：QR码可以使用不同的数据编码（有关字符编码模式的详细信息，请参见第3.2.2节）。它们的复杂性会影响可以存储在代码中的实际字符的数量。 例如，具有最低错误校正级别的QR码版本2最多可以保存77个数字字符，但只能包含10个汉字字符。

**3. QR码的安全**

**3.1 威胁模式**

人们可以区分两种不同的威胁模型来处理QR码。首先，攻击者可能会反转任何模块，将其从黑色改为白色或以其他方式进行更改。第二，更受限制的攻击者只能将白色模块更改为黑色，反之亦然。

**3.1.1 两种颜色**

攻击现有QR码的最简单的方法就是生成一个包含二维码的贴图，其中带有QR码的操作与原始QR码相同，并将其放在广告上的代码上。当然，这将需要一些准备工作，或者移动打印机和移动设备的设计应用程序。 至少当对一个选定的目标进行大规模的攻击时，准备所需的时间不应该是严重的限制。

由于这次攻击是微不足道的，我们决定将其排除在本文的范围之外。 但是，我们认为，使用这种方法来攻击真实世界的广告是大规模攻击的可行选择。

**3.1.2 单一颜色**

在这种情况下，我们仅限于修改单一颜色。 这个限制的背景在于攻击者试图通过使用笔修改单个海报的方案（从而减少了将白色模块更改为黑色的可能修改）。 这个限制是本文进一步概述的攻击的基础。

**3.2 攻击不同部位**

由于QR码包含大量不同的信息，包括有关版本，掩码和源编码的元信息，因此存在可以单独或组合定位的几个不同区域。

**3.2.1 掩码**

掩码用于生成QR码，分布良好的黑白模块（接近50:50，分布在整个代码上）。 这增加了图像的对比度，从而帮助设备对其进行解码。根据标准，当产生QR码时，应用8个指定的掩码，并对每个结果进行评级。选择符合等级的最佳分布掩码（使用正确掩码的效果可以在图4中看到。之前的栏显示了应用任何掩码之前的白色和黑色元素的数量，在数字之后之后使用最好的掩蔽）。在给定的QR码中总是只使用一个掩码，它使用强BCH编码在版本的单独块中与版本一起编码。

在表2的条件中，我将模块的行位置指向其列的位置。 对于条件有效的每个模块，掩码为黑色，其余为白色。

针对掩码可以在整个数据和纠错部分中发生很大的变化，但这仍然是另外应用其他方法的有用依据。更改掩蔽时的一个问题是使用强纠错算法分开进行编码。

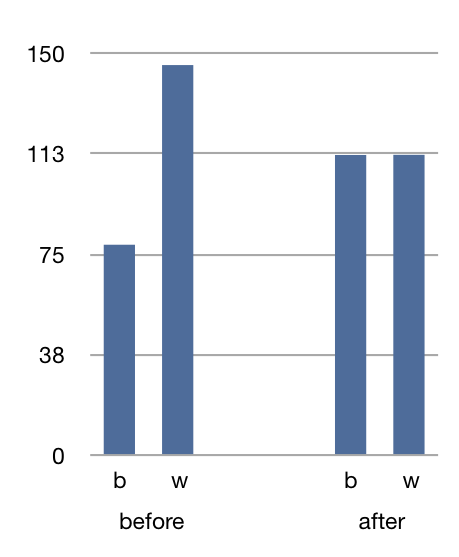


图4：屏蔽效果

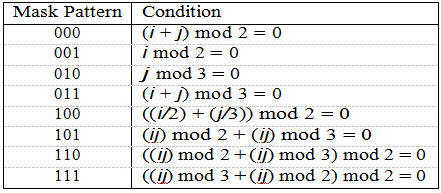


表2：掩码模式参考和条件

**3.2.2 字符编码（模式）**

为代码中包含的信息指定了几种不同的源代码（表3），从而最大限度地提高了复杂性的交换能力：

•数字模式（只需编码数字，从而可以在一张照片中打包大量数据），

•字母数字模式（一组包含大写字母的字符和数个额外的字符，如$或空白），

•8位模式（可根据JIS X 0201编码JIS 8位字符集（拉丁和假名））或

•汉字字符（根据JIS X 0208附件1的Shift JIS字符集移位编码表示）

来命名最受欢迎的。

字符编码本身在数据部分的起始位置由前导4位定义。 下表概述了模式的可能值：

改变模式指示器给编码数据带来全新的意义。 特别是当考虑8位字节模式而不是其他模式时，甚至是字母数字模式，而不是例如。数字模式，启动代码注入（例如SQL注入）可能变得可行（8位模式甚至允许使用控制字符更复杂的技巧）。

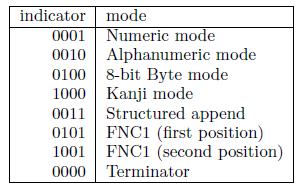


表3：模式指标

可以通过在一个QR码中混合不同的模式来安装对模式的高级攻击（请参见下面的部分）。

**3.2.3 字符计数指示器**

在模式指示符之后，下一位表示以下数据的字符数。 字符计数指示器的实际大小很大程度上取决于使用的模式和QR码的版本（更高版本包含更多数据，因此字符数量指示器更长）。 有关最受欢迎模式的长度，请参见下表（表4）。

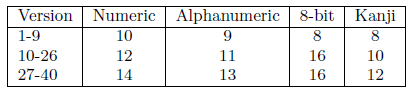


表4：字符计数指示符的长度

看这个目标，我们看到攻击的两个主要方法：产生缓冲区溢出或缓冲区下溢。

•缓冲区下溢：我们将字符计数指示符更改为与原始QR码相比较低的数字。因此，解码装置应仅将前几个字母解码为消息，而不排除其余部分。这在编码的原始链接包含后缀的情况下特别有用。由于数据部分的大小是固定的，所以在预期字节数之后的所有数据被看作是一个新的段（见混合模式），或者（在终端模式指示符的情况下）作为填充。由于这只是数据部分的一个微小的改变（只是长度被改变），所以这反过来也只会导致错误校正值的微小变化，并且仍然可以被解码。在使用混合模式的情况下，还应特别注意与其他目标的这种攻击的可能组合。

•缓冲区溢出：我们将字符计数器指示器更改为更高的数字，因此解码设备尝试将填充的部分解码为数据（如果我们甚至可以更改填充程序，我们获得有用的空间以包含其他数据）。同样，这种方法的缺点之一在于Reed-Solomon-Code的纠错（特别是错误检测）能力。尝试使用此方法启动代码注入攻击将是特别有趣的。

**3.2.4 混合模式**

可以在一个QR码内使用几种不同的模式（这对于在编码不同类型的数据时增加密度特别有用）。为了实现这一点，几个具有自己的模式指示器和特征数指示器的数据段只需要连接（见图5）。再次，我们可以确定利用此功能的QR码攻击的四种方法：

•更改段的模式：这样做就像对不使用混合模式的QR码的模式进行攻击，但是仅被减少到一个段，从而使数据的其他部分保持不变。

•插入新片段：将现有片段拆分成两个新片段，一个具有旧标题（模式指示符和字符计数指示符），另一个具有新定义的标题。

•删除现有段：使用附加数据覆盖模式指示符和段的字符计数指示符使用的空间，从而减少段数。此外，可以更改之前的段的字符计数指示符，以允许附加删除的段的数据。

•结构化追加：结构化追加模式允许连接多达16个QR码符号，因此可以在这个QR码序列中存储大量数据。它的设计方式是解码与读取符号的顺序无关。利用这一点的基本思想在于增加一个这样的一个段，指出另一个QR码就是原来的。然而，这种攻击需要提前准备贴纸，使其不太实用（这比起原始QR码顶部的贴纸更为重要，特别是由于结构化追加模式非常复杂）。

对模式的所有攻击（特别是插入和删除段）的主要问题是它们对纠错码字的影响通常很大。可能这种攻击可以与对纠错部分本身的攻击相结合使用。

应该尝试将现有细分的变化模式与数据部分（当然是错误校正部分）中的颜色变化相结合，因为如果没有足够的白色模块用于更改（记住作为先决条件，我们只考虑将一种颜色更改为另一种颜色，而无需进行相反操作）。同样将数字和字母数字的模式更改为8位会突然允许不可打印的控制字符（如删除）（图5）。

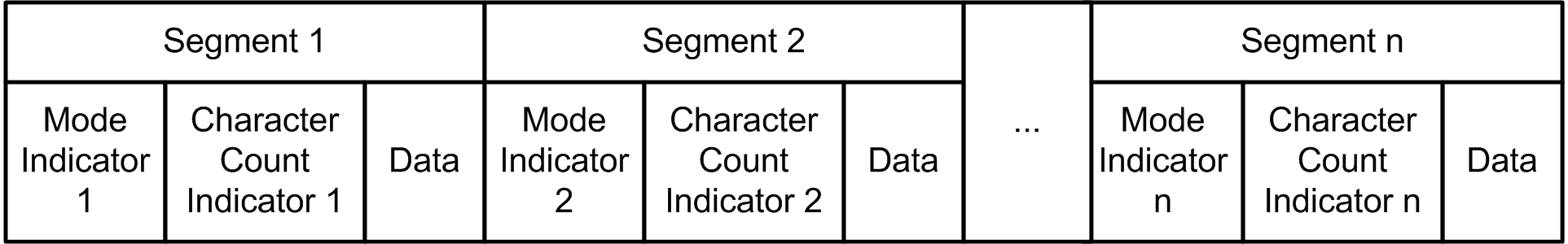


图5：包含多种模式的消息

**3.2.5 数据部分和纠错**

QR码的最大部分由包含数据和纠错码字的部分组成。数据部分本身可以由使用几种不同编码的段组成，每个段都有自己的标题，指定使用的模式和数据的长度（参见上面的模式）。对于给定版本和错误校正级别，可以容易地定义代表数据码字的QR码中的部分和表示纠错码字的部分，而无需解码，由于数据部分的长度不依赖于数据的实际长度（数据填充到全长的填充模式）。对于数据部分的位置，请参见第2节中的图3。数据部分的确切长度可以从标准中导出。

根据设计，底层数据的任何变化都直接反映在数据和纠错部分。 Reed-Solomon编码能够检测数据或纠错部分中的几个变化，即使在引入了中等数量的修改/错误之后，也可以对原始消息进行解码，即如果Qi表示消息Mi的100％正确的QR码，则与原始码Qi只有微小偏差的QR码Qr仍可被解码回到Mi。 这个特性虽然旨在保护代码的完整性尽可能地成为我们攻击的重要先决条件：我们不需要改变原始QR码Qi的含义，以完全匹配QR码Qj ，iƒ= j包含我们操纵的QR码Mj，我们只需要达到一个被解码成同一个消息的代码Qr。在传统的计算机安全方面，这类似于NOP雪橇。

图6说明了这种攻击的概况。绿色圆圈表示代表精确消息（即不包含错误）的QR码，蓝色圆圈由于Reed-Solomon码的纠错特征而被解码为相同消息的错误QR码组。 F表示通过编码检测为不正确但不能被纠正的QR码组。

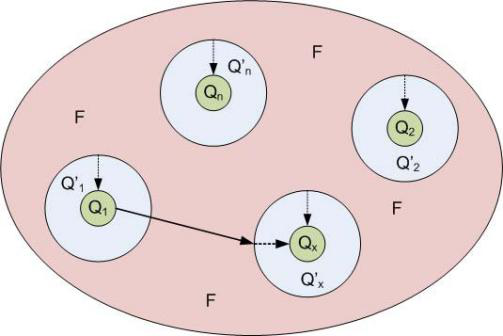


图6：攻击数据码字

1. **QR码作为攻击向量**

我们相信操纵的QR码可以用于大量的攻击。 根据读者是人类还是自动化程序（例如物流），不同的场景是可能的，并且在本部分中概述。

**4.1. 攻击自动进程**

由于QR码是编码信息的标准化方式，我们强烈认为，大多数软件开发人员将编码信息视为可能不安全的输入。如上一节所述，QR码的不同部分可以被操纵以便改变编码信息。 根据处理编码信息的程序，无论是物流，公共交通还是完全自动化的装配线，攻击读者软件以及后端在理论上都是可能的。没有适当的卫生，这可以由对手用于以下的非穷尽的攻击性列表。 使用RFID芯片和SQL注入的类似攻击已被证明是非常有效的[15]，因为这些示例中没有采用输入卫生。

•SQL注入：我们认为，许多自动化系统在数据库中存储和处理编码信息。通过向编码信息附加一个分号，后跟一个SQL查询（;drop table <tablename>），对后端数据库的操作是可能的（前提是DBMS允许单行中的多个查询）。这将删除命令中指定的表，导致拒绝服务攻击。更具体的攻击可能包括添加用户，执行系统命令（例如，通过使用Microsoft SQL Server上的存储过程xp\_cmdshell）或更改数据库中的价格或密码等数据。

•命令注入：如果将编码信息用作命令行参数而不进行整理，则可以轻易地利用它们代表攻击者运行任意命令，这可能会对操作系统的安全性造成灾难性的后果例如，在攻击者的控制下安装rootkit，DoS或将shell连接到远程计算机。

•欺诈：自动化系统的改变可以用于欺骗系统，例如欺骗系统来处理廉价产品A，同时处理更昂贵的产品B.

**4.2. 攻击人的互动**

人们无法读取代码，没有阅读器软件，存储在代码中的信息完全隐藏。 但是通过阅读操纵的QR码，阅读器软件或浏览器的可能性可能会被触发。

•网络钓鱼与网址嫁接：如果在增强现实场景中使用QR码用于链接，攻击者可能会设置一个假网站，并通过更改QR码重定向用户。如果需要某种形式的凭据来访问网站，这是很危险的。用户无法验证链接是否未被修改。

•欺诈：QR码通常用于广告中，以将目标受众指向特定优惠或有关特定产品的附加信息。如果QR码可以被操纵以将用户重定向到克隆的网站，对手可以在没有履行合同的情况下出售被请求的产品。受害者通过跟随链接相信广告公司。

•攻击读取器软件：如果编码信息未被消毒，则通过命令注入或传统缓冲区溢出可能会对计算机或手机上的读取器软件执行不同的实现。攻击者可能会控制整个智能手机，包括联系人信息或受害者的通信内容，如电子邮件或短信。

•社会工程攻击：根据这些攻击，根据攻击者的目标，启用更具体的攻击，如刺网钓鱼或其他社会工程变体。在附近的餐厅提供折扣的公司的停车场（而不是传统的USB驱动器攻击）上留下QR码的海报是可能会成功的新型攻击媒介。

**4.3. 进一步的研究**

本文提出的攻击的可能性为进一步研究开辟了一个相当大的领域。 主要目标在于对给定目标的一个或多个概述的攻击的准确分析和实际应用。 此外，应该调查QR码的哪些部分是最容易受到攻击的，以及可以采取什么对策来挫败本文提出的攻击。 在更一般的术语中，根据给定类型的攻击轮廓和关于黑/白分布，版本，掩蔽等特征，找到可用于测量QR码的脆弱性的度量值将是非常有趣的。

最后但并非最不重要的是，需要以相同的方式分析其他二维码如Aztec [12]或DataMatrix [10]，以识别可能的攻击向量并找到合适的对策。

1. **结论**

在本文中，我们概述了使用操纵QR码的可能攻击的危险。 由于QR码通过用于营销目的而越来越受欢迎，我们预计这种攻击将来将受到黑客社区越来越多的关注。 此外，目前许多移动设备（例如，智能电话）能够解码QR码并访问它们中包含的URL。这为信任主题增加了一个新的维度，特别是因为大多数用户在使用手机时不足够安全（这也使得可以使用新颖的网络钓鱼技术）。除了钓鱼之外，还可以使用QR码来执行许多其他攻击手段，无论是针对人类还是自动系统。如果在处理包含的数据之前没有执行适当的输入消毒，这一点尤其如此。

1. **引用**

[1] H. S. Al-Khalifa. Utilizing qr code and mobile phones for blinds and visually impaired people. In ICCHP, pages 1065–1069, 2008.

[2] A. Alapetite. Dynamic 2d-barcodes for multi-device web session migration including mobile phones.

Personal and Ubiquitous Computing, 14(1):45–52, 2010.

[3] R. Bose and D. Ray-Chaudhuri. On a class of error correcting binary group codes\*. Information and control, 3(1):68–79, 1960.

[4] M. Canadi, W. H¨opken, and M. Fuchs. Application of qr codes in online travel distribution. In ENTER, pages 137–148, 2010.

[5] J. Gao, V. Kulkarni, H. Ranavat, L. Chang, and

H. Mei. A 2d barcode-based mobile payment system. In MUE, pages 320–329, 2009.

[6] J. Z. Gao, L. Prakash, and R. Jagatesan. Understanding 2d-barcode technology and applications in m-commerce - design and implementation of a 2d barcode processing solution. In COMPSAC (2), pages 49–56, 2007.

[7] J. Z. Gao, H. Veeraragavathatham, S. Savanur, and

J. Xia. A 2d-barcode based mobile advertising solution. In SEKE, pages 466–472, 2009.

[8] A. Hocquenghem. Codes correcteurs daˆA˘ Z´erreurs.Chiffres, 2(147-156):4, 1959.

[9] Y.-P. Huang, Y.-T. Chang, and F. E. Sandnes. Ubiquitous information transfer across different platforms by qr codes. J. Mobile Multimedia, 6(1):3–14, 2010.

[10] ISO 16022:2006. Data Matrix bar code symbology specification. ISO, Geneva, Switzerland.

[11] ISO 18004:2006. QR Code bar code symbology specification. ISO, Geneva, Switzerland.

[12] ISO 24778:2008. Aztec Code bar code symbology specification. ISO, Geneva, Switzerland.

[13] S. Lisa and G. Piersantelli. Use of 2d barcode to access multimedia content and the web from a mobile handset. In GLOBECOM, pages 5594–5596, 2008.

[14] I. Reed and G. Solomon. Polynomial codes over certain finite fields. Journal of the Society for Industrial and Applied Mathematics, 8(2):300–304, 1960.

[15] M. R. Rieback, B. Crispo, and A. S. Tanenbaum. Is your cat infected with a computer virus? In PERCOM ’06: Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications, pages 169–179, Washington, DC, USA, 2006. IEEE Computer Society.